TRANSLATION TECHNOLOGY: BREAKING THE LANGUAGE BARRIER

On Sept. 10, 2001, the National Security Agency intercepted two Arabic-language messages. "The match is about to begin," said one. "Tomorrow is zero hour," said the other. Unfortunately the sentences were not translated until a day after the September 11 attacks. In fact, the phrases may have not been translated with such a quick turnaround had the horrific events not happened. In the aftermath of the September 11 attacks, it became clear that the United States continually faces serious technical and man-power hurdles to rapidly translate important intelligence.

Translation demands only increased with the onset of Operation Enduring Freedom and the subsequent insurgency. Decision-makers require rapid translations of broadcast news shows, blogs, and newspaper articles. These translations are critical to decision-makers on all levels so that they may properly gather intelligence and understand how people respond to U.S. actions. Further, soldiers who complete daily missions on city streets face serious communication barriers. Although human translation may be a key in solving the current problems, there is a serious shortage in the number of reliable human translators who can handle all the incoming information. Throw in the fact that there is not one "Arabic" – in and around Iraq, millions of people speak various dialects including Baghdadi Arabic and Moslawi Arabic – and it is a major challenge. And of course, the intelligence agencies still need – as they did on Sept. 10, 2001 – to sift through and translate intelligence coming from many sources. With these substantive challenges, translation technology is crucial to moving forward successfully.

DARPA has been a key sponsor of machine translation, as well as computer processing of language (speech and text) work for three decades. The goal is to create technology that will automatically translate spoken or written words from foreign languages into grammatically correct English. The technology is moving toward a direction that will allow the translations to happen in real time. These goals are set forth in ambitious research efforts and these efforts are poised to come close to achieving their goals in certain specific contexts with Modern Standard Arabic and Mandarin Chinese. The largest of these efforts is the five-year, multimillion-dollar-per-year Global Autonomous Language Exploitation (GALE) program, which seeks real-time translation of Modern Standard Arabic and Chinese print, Web, news, and television feeds. GALE has an ambitious goal of reaching 95 percent accuracy without human mediation. The second program is the Spoken Language Communication and Translation System for Tactical Use (TRANSTAC) program. TRANSTAC's goal is highly specific: a portable two-way speech translation system that enables an average soldier to communicate with a person who cannot speak English.

DoD photo

Photo courtesy of DARPA

Courtesy of Voxtec International Inc.

Courtesy of Integrated Wave Technologies, Inc.

Opposite: A U.S. Special Forces soldier uses a Phraselator device with the debriefing module to determine where enemies have gone, and where weapons and explosives are stored in Iraq during Operation Iraqi Freedom. Far left: The Voice Response Translator, a product of Integrated Wave Technologies and funded by DARPA, allows for hands-free, eyes-free operation; the user speaks into the headset in English, which is then translated into one of 125 languages and subsequently broadcast via a small speaker. Left: In 2001, DARPA helped to produce a one-way translation device called Phraselator for use by warfighters. The device uses several hundred pre-programmed phrases and offers speech output for Arabic, Pashto, and other languages. Above: An eTAP Arabic Deployment Unit is used at Strong Angel, a reoccurring event where military personnel and civilians meet to demonstrate technologies that can be of assistance in post-disaster or post-conflict areas.

The programs fund competing projects by giants in speech-recognition and language-processing research – BBN Technologies in Cambridge, Mass; IBM in White Plains, New York; and SRI International in Palo Alto, Calif., in the GALE program; the same three plus Sehda, Inc., of Mountain View, Calif., as part of the TRANSTAC program. And while both programs are still under way, they have produced early payoffs that suggest that real-time, highly accurate machine translation might well be achieved, in unlimited context for GALE and in some constrained contexts for TRANSTAC in the next few years.

GALE began in September 2005 and is planned to run until September 2010. As a baseline, the program began with a BBN translation system called eTAP – which was itself the fruit of earlier DARPA programs. Since 2004 this system has been installed at 12 locations, including Central Command in Tampa, Fla., and the 10th Special Forces Group in Balad, Iraq. The IBM starting point, called Translingual Automated Language Exploitation System, was deployed in 2005 at two sites in Hawaii. These systems were put to work translating television and Webcasts. The systems capture Arabic audio, convert the audio to text (this process is a by-product rather than a separate algorithm), and translate the native text into English. The systems also store the translations in a searchable database. English-language queries of that database produce relevant video segments as well as translations of the speech in English. These early deployed systems were only about 55 percent accurate in translating text documents, and 35 percent accurate at translating speech segments. As the quality was poor, the translations were only used to decide what was important enough to be translated by humans. The systems greatly reduced the human workload since only approximately 5 percent of the documents were determined to be important enough to require human translators.

The new goal, under GALE, is for the competing teams at BBN, IBM, and SRI to reach more difficult accuracy milestones each year, stepping up to 95 percent accuracy on 95 percent of text documents and 90 percent of speech documents by 2010. "We know that we will achieve machine translation of the caliber that will only necessitate final inspection by humans," said GALE Program Manager Dr. Joseph Olive. "It will take over from human translation work, period. And the search results will deliver to the military the information they need, when they want it." The GALE processing engines are using technology from rule-based and statistical translation methods. For many years, machine translation was based on rules written by human linguists for each language. "The problem with that paradigm is that as the rules grow, they grow more complicated. And since they are interdependent, adding one more rule becomes a nightmare. The second approach is to use statistical paradigms; the computer finds short segments in the text of the foreign language and retrieves corresponding

segments in English. The most probable segments are joined in the most probable order to generate the translated English sentence," said Olive.

The competing GALE performers are combining technology learned through both approaches. The performers add rules to statistical likelihoods to keep improving accuracy. And they are trying to teach computers not only to continually learn but even to generate rules. To gain these improvements from translation algorithms, the teams are training their computers with large volumes of past news texts and broadcast recordings in both Mandarin Chinese and Modern Standard Arabic. The machines analyze structure and content and develop a deeper understanding of all possible contexts, rules, and usages. (The task is especially difficult for Mandarin Chinese, which relies more heavily on context, and lacks pronouns and other linguistic signposts that are common in English.) However, there is additional work in syntax and shallow semantics to provide more accuracy.

Whereas a literal translation of Arabic might produce the English result: "Sent Iran an envoy," the improved technology, employing Arabic grammar rules, yields "Iran sent an envoy." The new technology also copes with new circumstances for which the software has not been specifically trained by basing decisions on equivalent circumstances previously "learned" by the software. In addition, the software identifies names, and makes some sense of the roles played by people mentioned in a passage. For example, while a literal Arabic translation may produce "talked to the Mubarak the prime minister," software that takes into account roles produces the more accurate "Mubarak talked to the prime minister."

Already, the GALE technology approach is yielding concrete results, exceeding the performance of pre-GALE technologies. Some GALE technologies are now achieving very high accuracy percentages from structured speech. The 2007 goals for Arabic text were just as ambitious. For structured text – that is, written news reports – the 2007 goal was to hit 80 percent accuracy on 90 percent of documents. For more informal or looser text, such as blogs written by average people, broadcast newcasts and conversations, the goals are slightly lower. None of the teams hit all four goals for Mandarin, but BBN did achieve all four goals for Arabic.

For 2008, the Arabic bar is higher: DARPA is seeking an accuracy jump of at least 5 percent for each situation – and a 10 percent gain in the case of structured broadcast news. If any of the teams hit the 2008 accuracy goals, it would surpass any Arabic machine-language technology in the world, said John Makhoul, an electrical engineer who heads BBN's effort. "If we can achieve [2008's] targets, I don't think anybody would be close to that, anywhere in the world," he said. "It will be much more advanced than anything anybody has been able to achieve, never mind the 90-95 percent."

But Olive is expecting near-perfection. "The GALE aim is to eliminate the human translator," Olive said. He predicts that the GALE technology is one of those DARPA projects that will wind up revolutionizing language systems throughout the military and industry. Among other things, a commercial version with such high accuracy could greatly open up the frontiers of Internet search, by suddenly making translated Arabic and Mandarin documents available with English search queries. Makhoul noted that machine translation "has always been difficult. DARPA decided to put a concentrated effort on the problem, and basically have some of the best teams in the world attack the problem, and do it with real milestones and real accuracy goals in mind," he said.

The expertise behind the project is staggering. For example, BBN alone has brought in subcontractors from the University of Southern California, University of Maryland, University of Pennsylvania, University of Colorado, Northeastern University, MIT, Cambridge University, and Edinburgh University, plus American software companies and a French research institute. SRI and IBM have their own bevy of in-house researchers and subcontractors attacking the problem. Taking into account all three teams, "I would venture to say a large number of the [machine translation] brains in the United States, and a few of the best ones abroad, are working on some aspect of this," Makhoul said.

Olive sketches the ultimate vision: "The primary goal is to be able to watch, in real time, Arabic and Chinese news stations, and see in closed captions or on the side of the screen the English translation. You actually watch the programs, and get immediate translation. That is very valuable for troops in Iraq and for PSYOP [psychological operations] to be able to execute missions and see what local reaction is to their missions. And other than that, to have the ability to connect to Web sites, translate those, and do searches of the material." As always, the most critical documents can be farmed out to a network of human translators for the best possible translation. When the program's goals are met, humans will only be necessary to do the final editing.

Whereas the GALE program is working with banks of servers to convert news broadcasts to English, the TRANSTAC program aims to give soldiers and Marines a system that they can carry to help with speech translation. The problem might seem like a simplified version of the one GALE is tackling, but there are key differences and special challenges. First, news broadcasts are usually in a formal version of spoken Arabic; the speaker also speaks clearly with few or no mistakes. That's clearly very different from what is spoken by an average person on the street; rather than being formal Arabic, he speaks a local dialect filled with idioms, slangs, and stuttering and other disfluencies. Second, soldiers cannot carry server banks on their backs; what they need is a portable gadget that can recognize and translate what is spoken. The amount of memory and computing power on such platforms constrains how language translation is accomplished in a tactical setting.

"There's a long history of DARPA supporting research in machine translation, but we got very energized soon after the 9/11 attacks," said Dr. Mari Maeda, the program manager for the TRANSTAC program.

DARPA in 2001 helped produce a one-way translation device – called Phraselator – for use by warfighters. If the user limited the spoken input to one of several hundred pre-programmed phrases, he could choose speech output for Arabic, Pashto, and other languages. The system ran on a customized, ruggedized PDA and could be used for translating English to target languages but not vice versa. In 2004, DARPA began working on the software to translate colloquial Iraqi Arabic that predominates in Baghdad and surrounding areas in central Iraq. Soldiers might seek information about insurgent activities, interview drivers at a checkpoint, attempt to assist people seeking compensation for damaged vehicles, or try to help a family with a sick or injured child. A soldier is given a rugged laptop or customized hardware integrated with a microphone and speaker. Because of the constrained computing resource, TRANSTAC

The TRANSTAC program has resulted in software, seen in use in this photo, that makes it possible for military personnel to communicate effectively with a local, non-English-speaking population to seek information about insurgent activities or to assist with a sick child.

Photo courtesy of BBN

software utilizes a vocabulary size on the order of tens of thousands of words and focuses on tactically relevant questions and answers such as, "What is in the trunk of your car?" or, "How many people live in this house?" The spoken Iraqi Arabic answers are translated and the user listens to the synthesized English output.

To develop the software, Maeda and her team started by interviewing soldiers and Marines returning from Iraq to see what kind of conversations they were having – the so-called "tactical dialogue" that is critical to building rapport and gathering intelligence, manning checkpoints, and doing business. After the scenario development phase, they brought Iraqi Arabic speakers together with soldiers and Marines into recording studios to simulate different interactions. They recorded hundreds of hours of speech, which were then carefully transcribed and translated. "It is fairly labor-intensive to collect the data. This is not like collecting broadcast news recordings. We have to bring together role-players and actually generate the data." The recordings and transcriptions were then used by the research teams to train and build the software components – a speech recognizer, a translation module, and a speech synthesizer – that were then integrated into the two-way speech translation system.

Concrete results have flowed from the effort. When the program got started, the first version took minutes to translate the uttered inputs. By 2007, the speed had improved so that the translated output begins to be generated almost as soon as the speaker finishes his utterance. Typically, a system can handle 25 questions and answers within 10 minutes. For hardware, the program is using a Toughbook tablet computer with a combined speaker-microphone plug-in. Customized hardware solutions are also being explored. Formal evaluations take place twice a year with military users and foreign speakers, and significant improvements in translation accuracy and speeds have been achieved since the inception of the program. The goal is that by 2009, the system will be good enough to use in limited tactical contexts.

Once the machine is deployed, the software will need to be updated and enhanced to recognize new words that were not anticipated previously. For example, a soldier will be able to enhance the vocabulary set with local terms that would have gone unrecognized by the original system or even English acronyms that are constantly being added into the military vocabulary. Finally, Maeda said, "our goal is that the software can be adapted to other languages as military needs require. We should have technology tools to be able to have a working translation system in a new language in less than 100 days." DARPA has recently tested the "quick-response" capability with another language, Farsi.

The TRANSTAC program is tackling a DARPA-hard problem that will have an enormous impact for future warfighters, humanitarian aid workers, and others who must surmount language-barrier problems in their work.